第52卷 第3期 山 东 大 学 学 报 (理 学 版) 2017年3月
Vol. 52 No. 3 Journal of Shandong University(Natural Science) Mar. 2017

文章编号:1671-9352(2017)03-0074-08 DOI:10.6040/j.issn.1671-9352.2.2016.001

基于分形与自适应数据融合的 P2P botnet 检测方法

宋元章,李洪雨,陈媛,王俊杰
(中国科学院长春光学精密机械与物理研究所, 吉林 长春 130033)

摘要:提出了一种基于分形与自适应数据融合的 P2P 僵尸网络检测方法。构建单分形特性、多分形特性检测传感器,利用大时间尺度下的自相似性和小时间尺度下的局部奇异性刻画网络流量特征,利用 Kalman 滤波器检测上述特性是否异常。为获得更精确的检测结果,提出了一种自适应数据融合方法,根据证据冲突程度自适应得选择 DST(Dempster-Shafer Theory)、DSmT(Dezert-Smarandache Theory)对上述检测结果进行融合。而且,考虑到了 P2P 应用对检测的影响。实验结果表明该方法检测准确度较高。

关键词:P2P 僵尸网络;自适应数据融合;Dempster-Shafer 理论;Dezert-Smarandache 理论
中图分类号:TP393.08 **文献标志码**:A
引用格式:宋元章,李洪雨,陈媛,等.基于分形与自适应数据融合的 P2P botnet 检测方法[J].山东大学学报(理学版),2017,52(3):74-81.

P2P botnet detection method based on fractal and adaptive data fusion

SONG Yuan-zhang, LI Hong-yu, CHEN Yuan, WANG Jun-jie
(Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, Jilin, China)

Abstract: A novel P2P botnet detection algorithm based on fractal and adaptive data fusion was proposed. Firstly, it built the single-fractal detection sensor and the multi-fractal detection sensor, and they used the self-similarity under the large time scale and the local singularity under the small time scale to describe the characteristics of network. Kalman filter was used to detect abnormalities of the above characteristics. To get the more accurate detection result, an adaptive data fusion method based on DST (Dempster-Shafer Theory) and DSmT (Dezert-Smarandache Theory) was proposed. Depending on the conflict factor of evidences, DST and DSmT were adaptively utilized to fuse the results of two above detection sensors to get the final result. The side effects on detecting P2P botnet which P2P programs generated are considered. The experiments show that the proposed algorithm is able to detect P2P botnet with high accuracy.

Key words: P2P botnet; Adaptive Data Fusion; Dempster-Shafer Theory; Dezert-Smarandache Theory

0 引言

僵尸网络(botnet)是一种由大量恶意主机组成的网络,攻击者(botmaster)可以通过僵尸网络的二次注入过程对主机节点的负载进行重注,这样可以较容易地更改攻击类型。随着构建僵尸网络技术的发展,基于 P2P 网络的非集中式结构被用来构建新型僵尸网络,这种结构是分散的,没有集中的控制中心,可有效避免针对单点失效的抑制手段,具有更高的可靠性、健壮性。P2P 僵尸网络检测是当前网络和信息安全领域的热点,相关研究分析如下。

王志等[1]在对 bot 程序执行轨迹进行分析的基础上,提出了一种发掘僵尸网络控制命令集合的方法,对 bot 程序覆盖率特征进行分析,获得其执行轨迹,进而实现僵尸网络控制命令空间的发掘。

臧天宁等[2]对已知僵尸网络内部通信行为进行特征提取,并利用这些特征定义云模型,进而分析判断已知 bot 主机群的隶属关系。

Holz 等[3]在深入分析Storm机理的基础上提

收稿日期:2016-08-18; 网络出版时间:2017-02-18 09:36
网络出版地址:http://www.cnki.net/kcms/detail/37.1389.N.20170218.0936.002.html
基金项目:国家 863 高技术研究发展计划资助项目(2011AA7031024G);国家自然科学基金资助项目(90204014)
作者简介:宋元章(1986—),男,硕士,助理研究员,研究方向为网络安全和分布式计算.E-mail:songyuanzhang@163.com

出了一种遏制僵尸网络的方法,通过发布虚假的 key 扰乱命令与控制机制(Command and Control, C&C)从而达到遏制僵尸网络规模扩大的目的。

在协同检测方面,王海龙等[4]提出的协同检测模型可以在信息、特性和决策 3 个层次进行协同,臧天宁等[5]提出的协同检测模型可以分析各种安全事件之间隐藏的关联关系,即使它们发送的地理位置不同、时间段不同。

文献[6-9]深入分析和总结了僵尸网络演化和发展进程,对检测、防御、遏制等方面深入地研究和展望,并对当前僵尸网络相关研究提出了相应建议。

综上所述,当前 P2P 僵尸网络检测研究以下几个问题亟需分析和研究:

(1) 大部分方法主要关注于 P2P 僵尸网络特有的特性。当出现新型的网络结构、网络协议和攻击负载类型时,这些方法将不再使用,漏报率较高;

(2) 大部分方法未考虑网络场景中正常运行的 P2P 应用对检测的影响。P2P 僵尸网络和 P2P 应用具有比较相似的的流量特征,本质上 P2P 僵尸网络是一种“恶意”的 P2P 网络,若忽略了网络场景中正常运行的 P2P 应用对检测的影响,则误报率较高;

(3) 大部分方法进行僵尸网络检测时使用数据挖掘、机器学习等方法,需要使用大量历史数据和先验知识对分类器事先训练,检测效率不理想。

本文提出一种基于分形与自适应数据融合的 P2P 僵尸网络检测方法:

(1) 本文方法主要关注其“共有”异常,从网络流量的内在特性出发,将网络流量看作信号进行处理,利用分形理论综合分析网络流量在不同时间尺度、不同视角下的特性,利用这些特性的异常来检测僵尸网络,因为上述特性不依赖于特定类型的僵尸网络,所以当出现新型的网络结构、网络协议和攻击负载类型时,本文方法仍然能进行有效检测;

(2) 本文方法详尽考虑了网络场景中正常运行的 P2P 应用对检测的影响;

(3) 本文方法利用自适应数据融合方法对检测结果进行决策级数据融合,无需大量历史数据、先验知识,并可通过积累证据缩小假设集。

1　P2P 僵尸网络检测方法

1.1　检测方法概述

基于分形与自适应数据融合检测 P2P 僵尸网络的处理过程,见图 1。

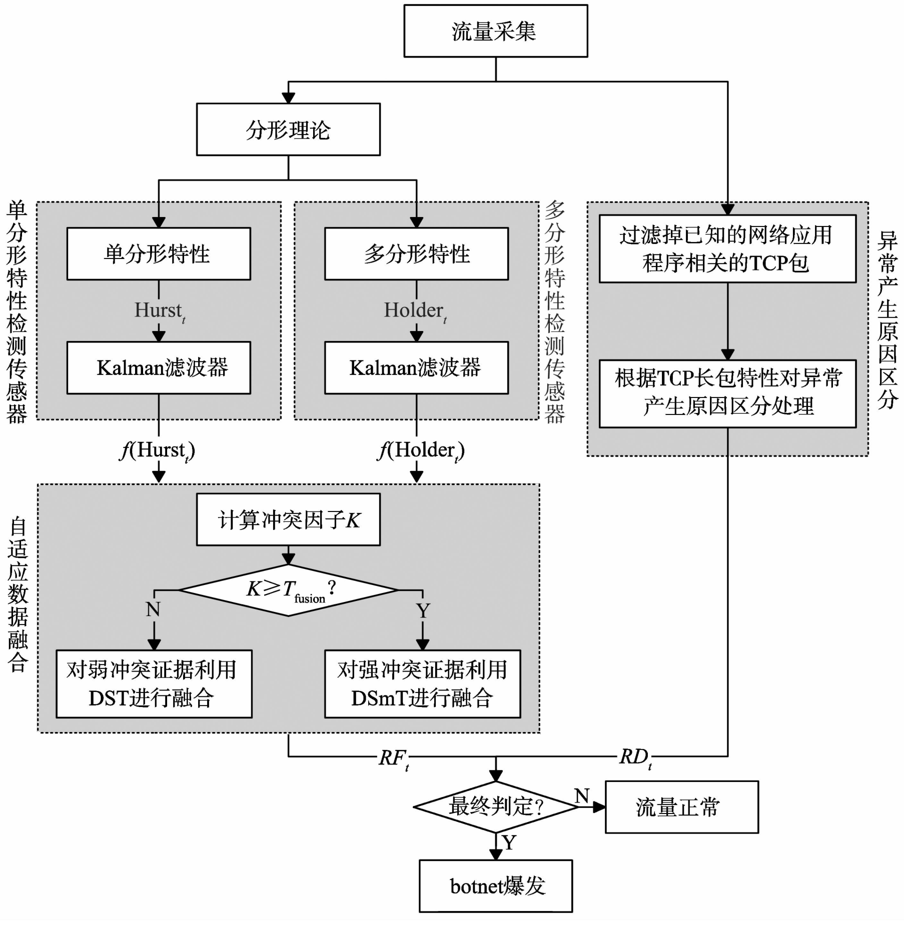

图 1　本文检测方法

Fig. 1　Process of the detection method

（1）构建单分形特性检测传感器和多分形特性检测传感器,用来检测网络流量在大时间尺度下的自相似性和小时间尺度下的局部奇异性是否异常;

（2）为获得更精确的决策级数据融合结果,在分析现有方法的基础上提出了一种自适应数据融合方法,根据证据冲突程度的不同自适应得选择DST、DSmT对上述检测传感器的检测结果进行决策级数据融合:对于弱冲突证据利用DST进行融合,对于强弱冲突证据利用DSmT进行融合;

（3）鉴于P2P僵尸网络和P2P应用流量特征相似,利用TCP包的特征对异常产生原因区分;

（4）获得最终检测结果。

1.2　构建流量异常检测传感器

分形通常指一个粗糙或零碎的几何形状,可以分成若干个部分,并且每一部分都(至少近似地)是整体缩小后的形状。分形一般具有以下特征[10]:

（1）在任意小尺度上都有精细结构;

（2）太不规则,以至于无论是整体还是局部都难以用传统欧式几何来描述;

（3）具有近似的或统计的自相似形式。

研究表明网络流量存在分形特性,具体表现为大时间尺度下的自相似性(单分形特性)和小时间尺度下的局部奇异性(多分形特性)[11]。经分析可知,P2P僵尸网络爆发时会导致IP包大量增加,使得大时间尺度下自相似性和小时间尺度下的局部奇异性发生变化,因此可构建相应流量异常检测传感器,根据上述特性刻画网络流量特征并通过利用Kalman滤波器检测是否存在异常。

1.2.1　单分形特性检测传感器

单分形特性检测传感器用来检测网络流量在大时间尺度下的自相似性是否异常。单分形特性(Single-fractal)体现的是网络流量在大时间尺度下的自相似性[12]。若一个连续时间随机过程$X(t)$满足

$$X(at)=a^{H}X(t)\quad \forall a>0, \tag{1}$$

则称$X(t)$具有自相似性。其中,参数H称为Hurst指数,描述过程的自相似性,$0.5\leqslant H\leqslant 1$, Hurst指数值越接近0.5,自相似程度越低。经分析可知,P2P僵尸网络爆发时会导致网络流量的自相似性减弱,从而导致Hurst指数的值减小。针对Hurst指数估算方法分析发现[13],R/S法(Rescaled Range)受噪声等因素影响较小且计算量较小,所以本文采用该方法计算Hurst指数。

1.2.2　多分形特性检测传感器

多分形特性检测传感器用来检测网络流量在小时间尺度下的局部奇异性是否异常。一个维数无法描述非均匀分形过程的全部特征,应采用多重分形测度或维数的连续谱进行描述。Riedi等[14]对TCP流量分析发现:自相似性只是流量分形特性的一个方面,在较小时间尺度上流量表现出更为复杂的变化规律,特别是局部奇异性和突发性。多分形特性(Multi-fractal)延伸和细化了网络流量的自相似性行为,可以更灵活地描述局部时间内的不规则现象。

若一个连续时间随机过程$X(t)$满足

$$X(at)=a^{H(t)}X(t)\quad \forall a>0, \tag{2}$$

则称一个连续时间随机过程$X(t)$是多重分形的。其中,$H(t)$称为Holder指数,描述过程的局部奇异性和突发性。若Holder指数<1,则表示时间随机过程在某点周围的小区间内的所有尺度上都有较高程度的突发;若Holder指数>1,则表示时间随机过程变化较平缓,突发不明显。经分析可知,P2P僵尸网络爆发时会导致网络流量的局部奇异性增强,从而导致Holder指数的值减小。Holder指数的计算方法见式(3)[15]。对于某一随机过程$X(t)$,表示到t时刻为止网络中IP包的数目,将$X(0),\cdots,X(t)$分配到若干子区间中,子区间的长度为d,则

$$\mathrm{Holder}_t=\lim_{d\to 0}\frac{\log\left(\left|X\left(t+\frac{d}{2}\right)-X\left(t-\frac{d}{2}\right)\right|\right)}{\log(d)}。\tag{3}$$

1.2.3　利用Kalman滤波器检测异常

Kalman滤波器是信号处理领域中使用最广泛的时间序列预测方法,由时间更新方程和测量更新方程组成[16-20]。

（1）时间更新方程

$$\boldsymbol{X}_{t|t-1}=\boldsymbol{A}\boldsymbol{X}_{t-1|t-1}+\boldsymbol{B}\boldsymbol{U}_{t-1}, \tag{4}$$

式(4)中$\boldsymbol{U}_{t-1}$为$t-1$时刻系统的控制量,$\boldsymbol{A}$和$\boldsymbol{B}$为系统参数,$\boldsymbol{X}_{t-1|t-1}$为$t-1$时刻的后验状态估计,$\boldsymbol{X}_{t|t-1}$为t时刻的先验状态估计。

$$\boldsymbol{P}_{t|t-1}=\boldsymbol{A}\boldsymbol{P}_{t-1|t-1}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{Q}, \tag{5}$$

式(5)中$\boldsymbol{P}_{t-1|t-1}$为$\boldsymbol{X}_{t-1|t-1}$的后验估计误差协方差,$\boldsymbol{P}_{t|t-1}$为$\boldsymbol{X}_{t|t-1}$的先验估计误差协方差,$\boldsymbol{Q}$是噪声协方差。

（2）测量更新方程

$$\boldsymbol{X}_{t|t}=\boldsymbol{X}_{t|t-1}+\frac{\boldsymbol{P}_{t|t-1}\boldsymbol{H}^{\mathrm{T}}}{\boldsymbol{H}\boldsymbol{P}_{t|t-1}\boldsymbol{H}^{\mathrm{T}}+\boldsymbol{R}}(\boldsymbol{Z}_t-\boldsymbol{H}\boldsymbol{X}_{t|t-1}), \tag{6}$$

$$\boldsymbol{P}_{t|t}=\left(\boldsymbol{I}-\frac{\boldsymbol{P}_{t|t-1}\boldsymbol{H}^{\mathrm{T}}}{\boldsymbol{H}\boldsymbol{P}_{t|t-1}\boldsymbol{H}^{\mathrm{T}}+\boldsymbol{R}}\boldsymbol{H}\right)P_{t|t-1}, \tag{7}$$

式(6)中$\boldsymbol{Z}_k$为测量值,$\boldsymbol{H}$为测量系统的参数,$\boldsymbol{R}$为

测量噪声协方差,式(7)中 $\boldsymbol{I}$ 为单位矩阵。将表征自相似性的 Hurst 指数 Hurst_t、表征局部奇异性的 Holder 指数 Holder_t 作为测量值,利用 Kalman 滤波器得到相应的后验状态估计 R_{Hurst_t}、R_{Holder_t}。

1.3 自适应数据融合

决策级数据融合方法主要有贝叶斯方法和 Dempster-Shafer 理论(DST),其中被广泛使用的方法为 DST,但是其仍有许多局限性,Dezert-Smarandache 理论(DSmT)是近来备受关注的静态融合和动态融合方法,它虽然弥补了 DST 的局限性,但其计算量较大。在分析 DST 和 DSmT 的基础上,本文提出了一种自适应数据融合方法,将检测结果 R_{Hurst_k}、R_{Holder_k} 进行决策级数据融合后得到检测结果 RF_k。

1.3.1 DST

设 U 为随机变量 X 取值的论域,若 U 内所有元素互不相容称,则 U 为 DST 中随机变量 X 的识别框架。设识别框架为 $U=\{\theta_1,\theta_2,\cdots,\theta_n\}$,$2^U$ 为 U 的幂集,若对于函数 $m:2^U\to[0,1]$ 满足如下条件:

$$(1)\ m(\varnothing)=0; \tag{8}$$

$$(2)\ \sum_{A\in 2^U} m(A)=1, \tag{9}$$

则称 $m(A)$ 为 A 的基本信度赋值函数。

设识别框架 U 上有 2 个相互独立的证据,与之相应的基本信度赋值为 m_1 和 m_2,与之相应的焦元为 $A_1,\cdots,A_k$ 和 $B_1,\cdots,B_r$,则 Dempster 组合规则见式(10),其中 K 为冲突因子[21]。

$$m(X)=\begin{cases}\dfrac{\sum\limits_{\substack{A_i,B_j\in 2^U\\ A_i\cap B_j=X}} m_1(A_i)m_2(B_j)}{1-K} & X\neq\varnothing,\\ 0 & X=\varnothing;\end{cases} \tag{10}$$

$$K=\sum_{\substack{A_i,B_j\in 2^U\\ A_i\cap B_j=\varnothing}} m_1(A_i)m_2(B_j)。 \tag{11}$$

1.3.2 DSmT

对于 DST,当直接使用 Dempster 组合规则对强冲突(冲突因子 K 趋近 1)的证据进行融合时在某些情况下会出现有悖常理的结果[22]。为解决上述问题,当前主要有两种方法:第一种方法为在 DST 框架下对 Dempster 组合规则进行改进,例如 Murphy 组合规则、Smets 组合规则、Yager 组合规则、Dubois-Prade 组合规则等;第二种方法为提出全新的组合规则,例如 DSmT。它对证据的冲突项进行了保留,将它们作为信息融合的焦元,并对冲突进行了重新地分配。另一方面,DST 要求识别框架中的元素互不相容,而 DSmT 不要求识别框架中的元素互不相容,它引入了不确定性的概念,因此当识别框架中元素之间的界限模糊、不确定、不精确且难以细分时,DsmT 亦可充分发挥其优势。

总之,DSmT 是 DST 的一种扩展,相比于 DST 幂集 2^U,DSmT 是基于 Dedekind 格子模型 D^U 建立的识别框架的超幂集作为数据融合空间。设识别框架 U,D^U 为 $U=\{\theta_1,\theta_2,\cdots,\theta_n\}$ 的超幂集,若对于函数 $m:D^U\to[0,1]$ 满足如下条件:

$$(1)\ m(\varnothing)=0; \tag{12}$$

$$(2)\ \sum_{A\in D^U} m(A)=1; \tag{13}$$

则称 $m(A)$ 为 A 的广义基本信度赋值函数。

当处理某些需要考虑已知约束的融合问题时使用混合 DSm 组合规则,设识别框架 U 上有 $n(n\geqslant 2)$ 个相互独立的证据,混合 DSm 组合规则见式(14)。

$$m(X)=\delta(A)[S_1(A)+S_2(A)+S_3(A)], \tag{14}$$

$$\delta(A)=\begin{cases}0, & A\ \text{为空集或由于约束条件而强制转换为空集的集合},\\ 1, & A\notin\varnothing,\end{cases} \tag{15}$$

$$S_1(A)=\sum_{\substack{X_1,X_2,\cdots,X_k\in D^U\\ X_1\cap X_2\cap\cdots\cap X_k=A}}\prod_{i=1}^{k} m_i(X_i), \tag{16}$$

$$S_2(A)=\sum_{\substack{X_1,X_2,\cdots,X_k\in\varnothing\\ [u(X_1)\cup\cdots\cup u(X_k)=A]\vee[(u(X_1)\cup\cdots\cup u(X_k)\in\varnothing)\wedge(A=\theta_1\cup\theta_2\cup\cdots\cup\theta_n)]}}\prod_{i=1}^{k} m_i(X_i), \tag{17}$$

$$S_3(A)=\sum_{\substack{X_1,X_2,\cdots,X_k\in D^U\\ (X_1\cup X_2\cup\cdots\cup X_k)=A\\ X_1\cap X_2\cap\cdots\cap X_k=\varnothing}}\prod_{i=1}^{k} m_i(X_i)。 \tag{18}$$

其中,$\delta(A)$ 为集合 A 的特征非空函数,$S_1(A)$ 表示基于自由 DSm 模型的 k 个相互独立证据的经典 DSm 组合规则,$S_2(A)$ 表示将所有相对和绝对的空集的信度质量传递给总的或相对的未知集,$S_3(A)$ 表示将相对于空集的信度质量之和传递给非空集。

1.3.3 DST 与 DSmT 自适应数据融合方法

在弱冲突情况下利用 DST 进行融合是普遍认为非常有效的方法[23],在强冲突、信息模糊情况下利用 DSmT 进行融合可以有效解决冲突分配的问题,获得比 DST 更优的融合结果,但 DSmT 计算量和存储量大,且在弱冲突情况下效果不如 DST。本文提出了一种自适应数据融合方法,根据冲突程度自适应得选择 DST、DSmT 进行融合,见图 2。

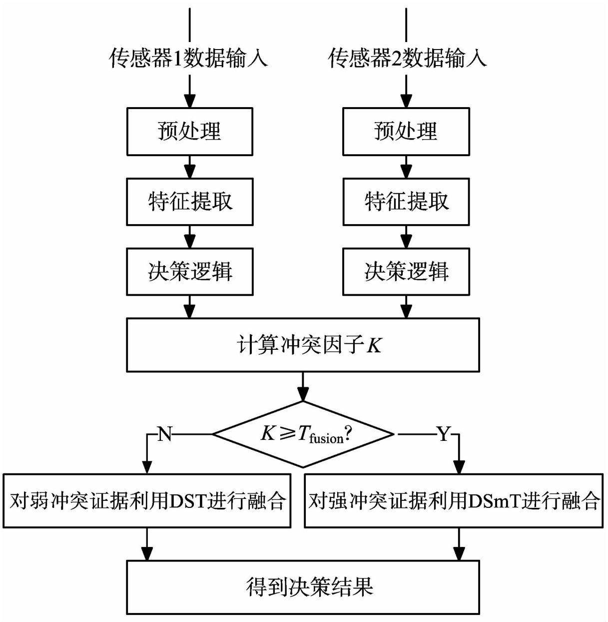

图 2　DST 与 DSmT 自适应数据融合

Fig. 2　Adaptive Data Fusion of DST and DSmT

设识别框架 $U=\{\theta_1,\theta_2,\cdots,\theta_n\}$：

（1）根据式(11)计算冲突因子 K，并设定阈值 T_{fusion}；

（2）若 $K\geqslant T_{\text{fusion}}$（强冲突），则根据式(14)利用 DSmT 进行融合。否则，根据式(10)利用 DST 进行融合。

经分析可知，对于不同类型的证据，阈值应不同，甚至可能是多个变化的点或区间。针对该问题，本文不再深入探讨，为简化计算量，设 $T_{\text{fusion}}=0.7$[24]。

1.4　异常产生原因区分

从网络结构、流量构成的角度上分析，P2P 僵尸网络是一种“恶意”的 P2P 网络，因此上述异常也可能是正常 P2P 应用的运行引起的。分析 P2P 应用的机制和 P2P 僵尸网络的生命周期可知：P2P 应用通常利用长度超过 1 300 字节的 TCP 进行数据传输，P2P 僵尸网络通常利用 TCP 包进行二次注入，两者的 TCP 包的长度有较大差异。设 TCP 包的数量为 N，长度超过 1 300 字节的 TCP 的数量为 N_L，利用 TCP 长包的比例 Pr 实现异常产生原因的区分，具体过程见图 3，其中 DPI（Deep Packet Inspection）详见文献[25]。设当前时刻为 t，定义函数

$$RD_t=\begin{cases}1 & Pr<T_{\text{TCP}}\\ 0 & Pr\geqslant T_{\text{TCP}}\end{cases}。\tag{19}$$

当 $Pr<T_{\text{TCP}}$时，上述流量异常是由 P2P 僵尸网络导致的可能性更大。可根据网络场景的不同利用 Kaufman 算法[26]对 T_{TCP}进行不同设定。

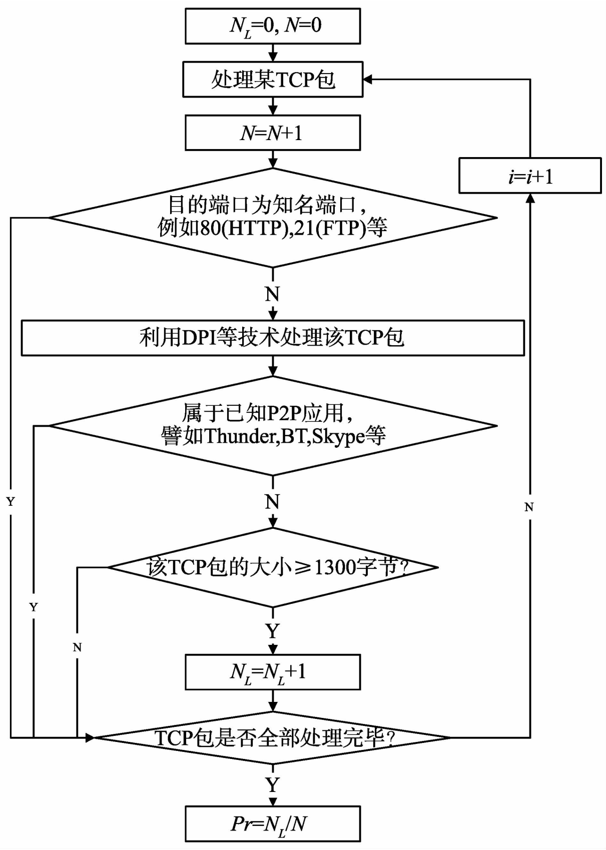

图 3　TCP 包处理过程

Fig. 3　Process of TCP Flow

1.5　检测方法处理过程

设当前时刻为 t，基于分形与自适应数据融合检测 P2P 僵尸网络的处理过程为：

（1）分析网络流量的分形特性

① 利用单分形特性检测传感器检测网络流量在大时间尺度下的自相似性是否存在异常：利用基于滑动窗口的估算 Hurst 指数的方法得到 Hurst_t，将 Hurst_t 作为 Kalman 滤波器的测量值检测网络流量自相似性特征的异常，得到检测结果 R_{Hurst_t}；

② 利用多分形特性检测传感器检测网络流量在小时间尺度下的局部奇异性是否存在异常：估算 Holder 指数 Holder_t，将 Holder_t 作为 Kalman 滤波器的测量值检测网络流量局部奇异性特征的异常，得到检测结果 R_{Holder_t}；

（2）利用自适应数据融合进行决策级数据融合，得到检测结果 RF_t；

（3）计算得到 TCP 流的 RD_t 值，以在一定程度上减弱 P2P 应用对检测的误差影响；

（4）利用加权平均法获得最终结果。

$$R_t=\alpha_t*RF_t+\beta_t*RD_t,\quad \alpha_t+\beta_t=1。\tag{20}$$

设判决 P2P 僵尸网络爆发的阈值为 T_{desion}，若 $R_t\geqslant$

T_{desion},则 P2P 僵尸网络爆发,可根据网络场景的不同利用 Kaufman 算法[26]对 T_{desion}进行不同设定。

2　实验与分析

2.1　实验环境

实验数据由两部分组成:第一部分为正常场景的网络数据,来自于某信息中心网络服务器;第二部分为 P2P 僵尸网络流量数据,来自于采用虚拟机(Virtual Machine, VM)技术参照文献[27]建立的实验环境,见图 4。为仿真大规模网络使用场景,利用 VM 工具在物理计算机上安装若干个虚拟计算机,在作为关键节点的虚拟计算机上安装网络数据包分析工具进行数据包采集和分析。在实验开始一定时间后向某些虚拟机注入 Storm bot 程序。本文中 VM 工具采用 Virtualbox,优点为配置方便、占用资源少、迁移性强;网络封包分析工具采用 Wireshark,优点为协议支持全面、细节丰富、支持数据重组。

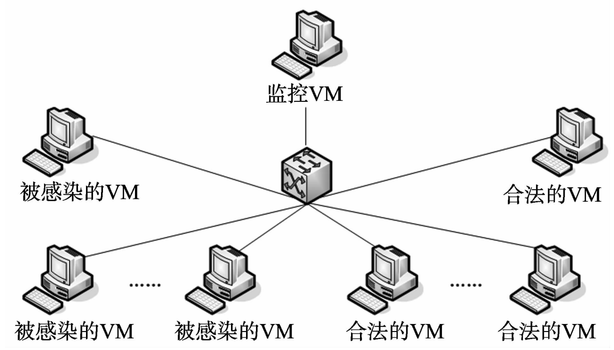

图 4　实验环境
Fig. 4　Experiment Environment

2.2　单分形特性实验

本实验主要观测单分形特性,正常网络场景中网络流量呈现较明显的自相似性。分析图 5 可得,在注入 Storm bot 程序后 Hurst 指数从 $t=420$ s 开始减小,说明自相似性开始减弱,表示网络流量出现异常。随着 P2P 僵尸网络规模逐渐扩大,网络流量呈现一种新的“病态”的自相似性行为,使得 Hurst 指数增大。

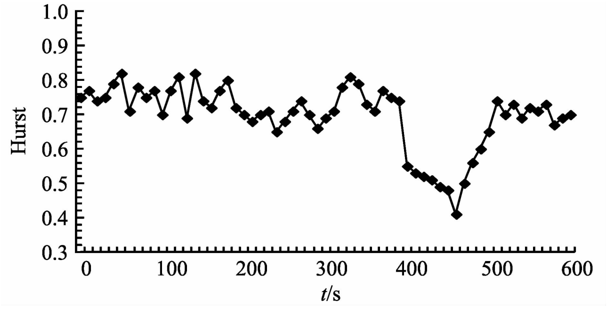

图 5　Hurst 指数
Fig. 5　Hurst exponent

2.3　多分形特性实验

本实验主要观测多分形特性,正常网络场景中网络流量呈现较明显的局部奇异性。分析图 6 可得,在一段时间注入 Storm bot 程序后,Holder 指数开始减小,表明网络流量局部奇异性程度增加,出现了异常。自相似性体现的是大时间尺度下的相关性,需要经过在 botnet 爆发后一段时间的数据积累才会发生变化,而局部奇异性体现的是小时间尺度下的突发性,在 botnet 爆发较短时间后就会发生变化,虽然 Holder 指数比 Hurst 指数敏感,但是容易造成误判。因此采用决策级数据融合的方法综合考虑自相似性和局部奇异性的变化,在降低漏报率的同时也降低误报率。

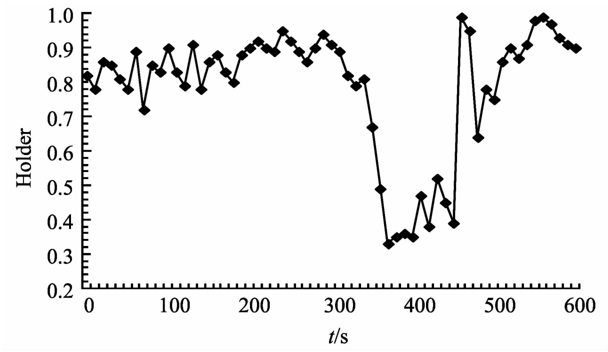

图 6　Holder 指数
Fig. 6　Holder exponent

2.4　检测准确度实验

为测试本文方法在不同情况下的检测准确度,本实验选择 4 种数据,并将本文方法与表 1 中方法进行对比,检测结果见表 2。

第 1 组实验数据来自某信息中心网络服务器,本文方法的检测结果与真实情况比较接近。

第 2 组实验数据来自某信息中心网络服务器中含有 P2P 应用的正常网络环境。比较第 1 组和第 2 组实验数据,发现对异常原因进行相应区分处理是必要的,本文方法可以有效地降低 P2P 应用对僵尸网络检测的影响,误报率相对较低。

将第 1 组实验数据与采集的 P2P 僵尸网络流量数据参照文献[28]中方法合并后获得第 3 组实验数据,本文方法的漏报率较低。

将第 2 组实验数据与采集的 P2P 僵尸网络流量数据参照相同方法合并后获得第 4 组实验数据,本文方法的检测结果比较理想。利用分形理论综合考虑多种网络内在特性来刻画网络变化的细节,同时观测网络流量在大时间尺度下的长相关性和在小时间尺度下的局部奇异性,通过利用 Kalman 滤波器检测上述特性的变化,并利用自适应数据融合方法进行决策级数据融合,同时考虑到了网络场景中

正常运行的 P2P 应用对检测的影响。其中,“1 018（873）”表示针对第 4 组实验数据,本文方法总共检测到了 1 018 次攻击、873 次是真正的攻击。

表 1 实验涉及检测方法概述

Table 1 Overview of detection method

序号	方法名称	方法概述
1	M-CUSUM[29]	通过多维 CUSUM 算法检测数据包数量的异常变化来检测僵尸网络。
2	KCFM[30]	抽取网络流量多个特征构成多维观测序列,利用 Kalman 滤波器检测流量的异常变化,利用 Multi-chart CUSUM 对每个维度的检测结果进行融合。
3	SF	仅利用单分形特性刻画网络流量特征,利用 Kalman 滤波器检测流量的异常变化。
4	MF	仅使用多分形特性刻画网络流量特征,利用 Kalman 滤波器检测流量的异常变化。
5	本文方法	利用大时间尺度下的自相似性和小时间尺度下的局部奇异性描述网络的内在特性,同时考虑到了网络场景中正常运行的 P2P 应用对检测的影响,利用自适应数据融合方法对检测结果进行合理有效地融合。

表 2 漏报率和误报率数据对比

Table 2 False negative rate and False positive rate

序号	实验数据说明	真实情况	M-CUSUM	KCFM	SF	MF	本文方法
1	正常	0	24	10	16	19	5
2	正常 + P2P	0	122	74	41	49	10
3	正常 + bot	1 000	764	758	732	893	885
4	正常 + P2P + bot	1 000	1 269（784）	1 247（823）	1 546（796）	1 693（815）	1 018（873）

3 结论

在分析 P2P 僵尸网络的典型代表 Storm 的生命周期和流量特征基础上,提出了一种基于分形与自适应数据融合的 P2P botnet 检测方法。利用分形理论构建 2 个检测传感器以检测网络流量在大时间尺度下的自相似性和小时间尺度下的局部奇异性是否异常。为取得更好的数据融合结果,在对贝叶斯方法、DST 和 DSmT 分析的基础上,提出了一种自适应数据融合方法,根据证据冲突程度不同自适应选择 DST 和 DSmT 对上述检测传感器的检测结果进行融合。同时,鉴于 P2P 僵尸网络和 P2P 网络的流量特征相似程度较高,利用 TCP 流量特征在一定程度上对引起异常的原因进行区分处理。下一步工作重点:如何更有效合理得对 DST 和 DSmT 进行结合和改进,如何更有效地消除 P2P 应用对检测的影响。

参考文献:

[1] 王志, 蔡亚运, 刘露, 等. 基于覆盖率分析的僵尸网络控制命令发掘方法[J]. 通信学报, 2014, 35(1):156-166.
WANG Zhi, CAI Yayun, LIU Lu, et al. Using coverage analysis to extract Botnet command-and-control protocol[J]. Journal on Communications, 2014, 35(1):156-166.

[2] 臧天宁, 云晓春, 张永铮, 等. 僵尸网络关系云模型分析算法[J]. 武汉大学学报·信息科学版, 2012, 37(2):247-251.
ZANG Tianning, YUN Xiaochun, ZHANG Yongzheng, et al. A botnet relationship analyzer based on cloud model[J]. Geomatics and Information Science of Wuhan University, 2012(37):247-251.

[3] HOLZ T, STEINER M, DAHL F. Measurements and mitigation of Peer-to-Peer-based botnets: a case study on storm worm[C]// 1st USENIX Workshop on Large-Scale Exploits and Emergent Threats San Francisco. [S.l.]: [s.n.], 2008: 3-12.

[4] 王海龙, 胡宁, 龚正虎. Bot_CODA: 僵尸网络协同检测体系结构[J]. 通信学报, 2009, 30(10A):15-22.
WANG Hailong, HU Ning, GONG Zhenghu. Bot_CODA: botnet collaborative detection architecture[J]. Journal on Communications, 2009, 30:15-22.

[5] 臧天宁, 云晓春, 张永铮, 等. 网络设备协同联动模型[J]. 计算机学报, 2011, 34(2):216-228.
ZANG Tianning, YUN Xiaochun, ZHANG Yongzheng, et al. A Model of network device coordinative run[J]. Journal of Computers, 2011, 34:216-228.

[6] 江健, 诸葛建伟, 段海新, 等. 僵尸网络机理与防御技术[J]. 软件学报, 2012, 23(1):82-96.
JIAN Jiang, ZHUGE Jianwei, DUAN Haixin, et al. Research on botnet mechanisms and defenses[J]. Journal of Software, 2012, 23:82-96.

[7] KARIM Ahmad, SALLEH Rosli Bin, SHIRAZ Muhammad, et al. Review: botnet detection techniques: review, future trends, and issues[J]. Journal of Zhejiang University-Science C (Computers & Electronics), 2014, 15(11): 943-983.

[8] JAIKUMAR Padmini, KAK Avinash C. A graph-theoretic framework for isolating botnets in a network[J]. Security and Communication Networks, 2015, 8(16):2605-2623.

[9] YAHYAZADEH Moosa, ABADI Mahdi. BotGrab: a negative reputation system for botnet detection[J]. Computers and Electrical Engineering, 2015, 41:68-85.

[10] KIM J S, KAHNG B, KIM D, et al. Self-similarity in fractal and non-fractal networks[J]. Journal of the Korean Physical Society, 2008, 52:350-356.

[11] GIORGI G, NARDUZZI C. A study of measurement-based traffic models for network diagnostics[C]// IEEE Instrumentation & Measurement Technology Conference. [S.l.]:[s.n.], 2007: 1-3.

[12] LELAND W E, TAQQU M S, WILLINGER W. On the self-similar nature of ethernet traffic (extended version)[J]. IEEE/ACM Trans on Networking, 1994, 2(1):1-15.

[13] KARAGIANNIS T, MOLLE M, FALOUTSOS M. Understanding the limitations of estimation methods for long-range dependence[R]. California: University of California, 2006: 11-15.

[14] RIEDI R H, VEHEL J L. Multifractal properties of TCP traffic: a numberical study[R]. Rocquencourt: INRIA, 1997: 6-17.

[15] MAULIK Krishanu, RESNICK Sidney. The self-similar and multifractal nature of a network traffic model[J]. Stochastic Models, 2003(19):549-577.

[16] 从爽, 孙光立, 邓科, 等. 陀螺稳定平台扰动的自抗扰及其滤波控制[J]. 光学精密工程, 2016, 24(1):169-177.
CONG Shuang, SUN Guangli, DENG Ke, et al. Active disturbance rejection and filter control of gyro-stabilized platform[J]. Optics and Precision Engineering, 2016, 24(1):169-177.

[17] 张百强, 储海荣, 孙婷婷, 等. 应用 RB 无迹卡尔曼滤波组合导航提高 GPS 重获信号后的导航精度[J]. 光学精密工程, 2016, 24(4):836-843.
ZHANG Baiqiang, CHU Hairong, SUN Tingting, et al. Precision improvement methodology for INS/GPS after GPS outage using RB-UKF[J]. Optics and Precision Engineering, 2016, 24(4):836-843.

[18] 陈东, 刘诗斌, 殷世民, 等. 光寻址电位传感器的噪声分析与信号处理方法研究[J]. 光学精密工程, 2016, 24(6):1456-1464.
CHEN Dong, LIU Shibin, YIN Shimin, et al. Research on noise analysis and signal processing method of light addressable potentiometric sensor[J]. Optics and Precision Engineering, 2016, 24(6):1456-1464.

[19] 刘志青, 李鹏程, 陈小卫, 等. 基于信息向量机的机载激光雷达点云数据分类[J]. 光学精密工程, 2016, 24(1):210-219.
LIU Zhiqing, LI Pengcheng, CHEN Xiaowei, et al. Classification of airborne LiDAR point cloud data based on information vector machine[J]. Optics and Precision Engineering, 2016, 24(1):210-219.

[20] 吴禄慎, 史皓良, 陈华伟. 基于特征信息分类的三维点数据去噪[J]. 光学精密工程, 2016, 24(6):1465-1473.
WU Lushen, SHI Haoliang, CHEN Huawei. Denoising of three-dimensional point data based on classification of feature information[J]. Optics and Precision Engineering, 2016, 24(6):1465-1473.

[21] YAGER Rr, LIU L. Classic works of the dempster-shafer theory of belief functions [M]. Berlin: Springer-Verlag, 2008: 23-49.

[22] MRUPHY C K. Combing belief function when evidence conflicts[J]. Decision Support System, 2000, 29(1):1-9.

[23] MATHON B R, OZBEK M M, PINDER G F. Dempster-shafer theory applied to uncertainty surrounding permeability[J]. Math Geosci, 2010, 42:293-307.

[24] SMARANDACHE F, DEZERT J. Advances and applications of DSmT for information fusion, Vol. 2[M]. Rehoboth:American Research Press, 2006: 15-39.

[25] SEN Subhabrata, SPATSCHECK Oliver, WANG Dongmei. Accurate, scalable in-network identification of P2P traffic using application signatures[C]// Proceedings of the 13th international conference on World Wide Web. New York: ACM, 2004: 512-521.

[26] KASERA S, PINHEIRO J, LOADER C. Fast and robust signaling overload control[C]// Proceedings of Ninth International Conference on Network Protocols. Riverside, USA: IEEE, 2001: 323-331.

[27] STEGGINK M, IDZIEJCZAK I. Detection Of Peer-To-Peer botnets [R/OL]. http://staff.science.uva.nl/~delaat/sne-2007-2008/p22/report.pdf.

[28] ZHAOA David, TRAOREA Issa, SAYED Bassam, et al. Botnet detection based on traffic behavior analysis and flow intervals[J]. Computers & Security, 2013(39):2-16.

[29] KANG Jian, ZHANG Jun-Yao, et al. Detecting new P2P botnet with multi-chart CUSUM[C]//International Conference on Networks Security, Wireless Communications and Trusted Computing. Wuhan: [s.n.] 2009, 1: 688-691.

[30] 康健, 宋元章. 利用多维观测序列的 KCFM 混合模型检测新型 P2P botnet[J]. 武汉大学学报(信息科学版), 2010, 35(5):520-523.
KANG Jian, SONG yuanzhang. Application KCFM to Detect New P2P Botnet Based on Multi-Observed Sequence[J]. Geomatics and Information Science of Wuhan University, 2010, 35(5):520-523.

(编辑:李晓红)